TRIEKB-SURG: TRIADIC EVOLUTIONARY KNOWLEDGE BRIDGING FOR TASK-HETEROGENEOUS FEDERATED SURGICAL VIDEO UNDERSTANDING

Anonymous authors
Paper under double-blind review

ABSTRACT

Federated learning (FL) enables institutions to collaboratively train surgical video understanding models without sharing raw data. However, most existing methods assume that all clients optimize the same task, which is unrealistic when institutions provide different forms of supervision. Although heterogeneous-task updates may share the same parameter space and thus be numerically aggregatable, they arise from distinct prediction spaces and optimization objectives. Therefore, parameter aggregatability does not guarantee knowledge transferability: a source update may be semantically irrelevant to the target, unreliable at the current training stage, or incompatible with the target objective. To address this challenge, we propose **Tri**adic **E**volutionary **K**nowledge **B**ridging (**TriEKB-Surg**), which frames task-heterogeneous FL as an evolving and client-specific knowledge-transfer problem. The triadic design considers semantic relevance, optimization-stage reliability, and task-compatible absorption. At the server, dual-prompt compatibility estimates cross-client relevance at coarse task-family and fine task-semantic levels, while an evolutionary curriculum combines source reliability with target-specific update agreement to construct personalized bridges. On the client side, cross-task alignment transforms transferred knowledge into task-compatible auxiliary supervision, while task-aware feature modulation mitigates optimization conflict. By determining *which* knowledge to transfer, *when* to trust it, and *how* to absorb it, TriEKB-Surg enables adaptive collaboration across heterogeneous surgical tasks. Experiments on five public surgical video datasets show that TriEKB-Surg achieves the best overall performance among the evaluated FL methods. Code will be made publicly available.

1 INTRODUCTION

Surgical video understanding is a key perceptual component of robot-assisted surgery (RAS), supporting downstream functions such as navigation, camera control, and autonomous intervention (Maier-Hein et al., 2022). Despite recent advances in segmentation and video foundation models, robust performance still depends heavily on annotated data and domain-specific adaptation (Kirillov et al., 2023; Yang et al., 2024). This dependence is particularly challenging in surgery, where videos are scarce, expensive to annotate, and difficult to share because of privacy and regulatory constraints (Maier-Hein et al., 2022). Federated learning (FL) offers a natural solution by enabling institutions to train models collaboratively without sharing their data (Sun et al., 2025; Yang et al., 2025b). However, surgical FL remains relatively underexplored, and most existing methods assume that all clients optimize the same task (Xu et al., 2025; Fang et al., 2026b). This assumption is restrictive in realistic collaborations, where institutions often annotate different tasks according to their clinical priorities, available expertise, and resources. Unifying these task spaces would require costly expert re-annotation and may not be practically feasible.

Task heterogeneity introduces a more fundamental challenge than conventional non-IID data. Standard FL assumes that client updates are functionally commensurable because they are induced by a shared prediction space and optimization objective. In task-heterogeneous FL, however, client updates may reside in the same parameter space while arising from distinct prediction spaces and optimization objectives. Consequently, ***parameter aggregatability does not imply knowledge trans-***

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

ferability. An update may be mathematically valid for aggregation yet semantically irrelevant to a target client, unreliable at the current training stage, or incompatible with the target objective. Moreover, cross-client transfer is both target-specific and time-varying. A source client may be useful to one target but not another, and its contribution may change as local models evolve during federated optimization. Even when a source is semantically related and locally reliable, its updates may still conflict with the target task or be difficult to absorb directly. The central question is therefore no longer merely how to aggregate client updates, but rather:

> *What determines whether heterogeneous knowledge can be transferred, and how should this transfer evolve during federated optimization?*

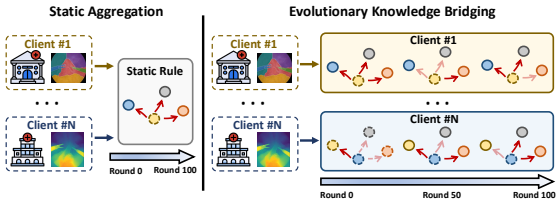


Figure 1: Comparison of vanilla static aggregation and the proposed evolutionary knowledge bridging.

We argue that effective cross-task collaboration depends on three complementary conditions. First, source knowledge should be semantically relevant to the target client. Second, the source should be sufficiently reliable at the current optimization stage. Third, the transferred knowledge should be represented in a form that the target task can safely absorb. None of these conditions is sufficient alone: semantic relevance does not guarantee reliable transfer, while even a relevant and reliable source may cause optimization conflict when trained under a different task objective. From this perspective, we formulate task-heterogeneous FL as an evolving, client-specific knowledge-transfer problem. Rather than following a fixed collaboration rule, each target client should dynamically determine which sources are relevant, when their knowledge can be trusted, and how to incorporate it into the local objective. This shifts federated collaboration from unconditional aggregation to adaptive, reliability-aware knowledge transfer, as illustrated in Fig. 1.

To this end, we propose **Tri**adic **E**volutionary **K**nowledge **B**ridging for federated surgical video understanding (**TriEKB-Surg**). On the server side, a dual-prompt compatibility module estimates cross-client relevance using coarse task-family and fine-grained task-semantic prompts, while an evolving curriculum combines round-dependent client reliability with target-specific update agreement. Together, these components construct personalized knowledge bridges that prioritize relevant and reliable sources while suppressing irrelevant, unstable, or conflicting updates. On the client side, the personalized model from the previous round serves as a teacher to generate complementary-task pseudo-labels, providing depth supervision for segmentation clients and segmentation supervision for depth-estimation clients. A cross-task alignment module converts this heterogeneous knowledge into task-compatible auxiliary supervision, while task-aware feature modulation combines dynamic conditions with fixed task-specific conditions to facilitate knowledge absorption and mitigate task-induced conflict. Through this design, TriEKB-Surg determines which knowledge to transfer, when to trust it, and how to incorporate it into the target task. Our main contributions are summarized as:

- We formulate task-heterogeneous FL as an evolving and client-specific knowledge-transfer problem, emphasizing that parameter aggregatability does not imply knowledge transferability.
- We propose a personalized evolutionary knowledge bridge that jointly models cross-client semantic relevance and round-dependent source reliability, and target-specific update agreement, enabling each client to dynamically select and weight useful source knowledge.
- We introduce a local cross-task alignment mechanism with task-aware feature modulation that converts heterogeneous transferred knowledge into task-compatible auxiliary supervision while mitigating task-induced optimization conflict.

2 Related Works

2.1 Surgical Video Understanding

Extensive studies have been conducted on surgical video segmentation and depth estimation, two dense prediction tasks that provide semantic and geometric cues for surgical video understand-

ing (Khan et al., 2025). Although their task-specific optimization objectives differ, both tasks require models to understand temporal and clinical information from surgical videos. Since both tasks require pixel-level prediction, existing methods commonly rely on encoder-decoder architectures to extract hierarchical visual features and recover dense outputs. To better model surgical video dynamics, many methods further incorporate spatial-temporal representation learning, optical-flow or motion cues, and geometric priors to capture local appearance and temporal consistency (Murali et al., 2016; Yue et al., 2024; Sellner et al., 2023; Li et al., 2025; Tao et al., 2023).

Recently, foundation models have reshaped this field (Ravi et al., 2025; Yang et al., 2024), with studies increasingly adapting foundation models to surgical video understanding through task-specific fine-tuning, aiming to inherit their robustness in video segmentation (Yue et al., 2024; Liu et al., 2024; Lou et al., 2025). However, these methods often assume that a single client or institution can collect sufficient surgical videos, or that data can be freely shared across institutions without considering privacy risks.

2.2 Federated Learning

FL enables distributed clients to collaboratively train models without sharing raw data, typically by exchanging model updates (McMahan et al., 2017). To mitigate client heterogeneity, various methods have been proposed from different perspectives, including constraining local optimization with proximal regularization Li et al. (2020), improving the optimizer (Wang et al., 2020), personalizing specific model components (Li et al., 2021; Arivazhagan et al., 2019), and introducing client-specific adaptation or aggregation mechanisms (Yang et al., 2025a). These methods mainly target data-level heterogeneity while assuming that clients optimize the same task objective (Charalampakis et al., 2026; Bian et al., 2025).

Recent studies attempt to explore the more challenging multi-task setting. MaTFL (Cai et al., 2023) estimates inter-client proximity from layer-wise weight updates and performs clustering-based selective aggregation to mitigate task-level heterogeneity. FedHCA2 (Lu et al., 2024) addresses heterogeneous-client federated multi-task learning by modeling client relationships and using hyper-conflict-averse aggregation, cross-attention aggregation, and learnable client-specific aggregation weights. Recently, HNFL (Fang et al., 2026a) explores multi-task FL for surgical video understanding by leveraging textual prompts to personalize aggregation weights. However, these methods still largely treat cross-client collaboration as a static learning process, where the collaboration pattern remains fixed across training. Such a static strategy may still introduce conflicts in task-heterogeneous FL, because the reliability and usefulness of different clients can change over time.

3 Methodology

3.1 Problem Statement

In conventional FL, clients are typically assumed to optimize the same task within a shared prediction space. Let K denote the number of clients, and let client $k \in \{1, \dots, K\}$ hold a private dataset $\mathcal{D}_k$. Its local objective is:

$$\mathbb{E}_{(x,y)\sim\mathcal{D}_k}\left[\ell\big(\mathcal{F}(x;\theta), y\big)\right], \tag{1}$$

where $\mathcal{F}(\cdot;\theta)$ denotes the shared model parameterized by θ, x and y denote an input sample and its corresponding label, and $\ell(\cdot)$ is the common task loss. Although client data distributions may differ, their updates remain functionally comparable because they are defined over the same prediction space and task objective.

In task-heterogeneous FL, clients may instead optimize different tasks. Let t_k denote the task assigned to client k. The corresponding local objective is

$$\mathbb{E}_{(x,y)\sim\mathcal{D}_k}\left[\ell_k\big(\mathcal{F}_k(x;\theta_k), y\big)\right], \tag{2}$$

where $\mathcal{F}_k(\cdot;\theta_k)$ denotes the client-specific model of client k, and $\ell_k(\cdot,\cdot)$ is the corresponding task loss. Although these models may share a common parameter space or backbone, their updates are induced by different prediction spaces and optimization objectives. Consequently, direct aggregation may combine semantically incompatible or conflicting updates, leading to negative transfer and degraded client performance.

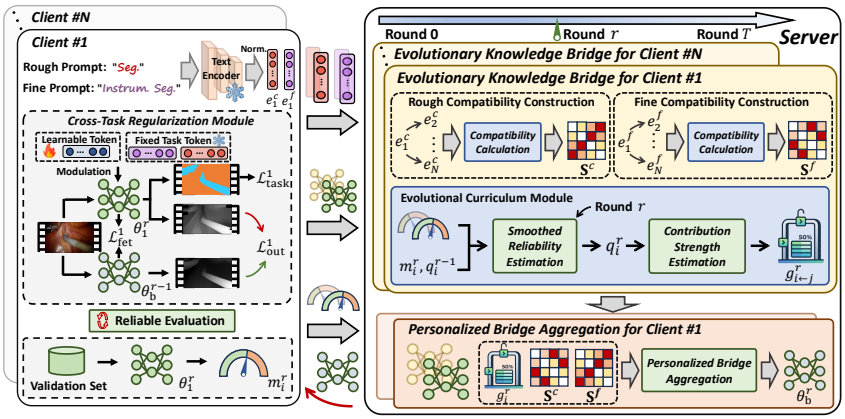


Figure 2: The proposed TriEKB-Surg framework.

3.2 Framework Overview

The proposed TriEKB framework is illustrated in Fig. 2. TriEKB-Surg addresses three fundamental questions in task-heterogeneous FL: which knowledge sources are transferable, when their contributions should be trusted, and how heterogeneous knowledge can be effectively absorbed by the target task. At the server side, a dual-prompt compatibility module models cross-client relevance using coarse task-family and fine task-semantic prompts. An evolutionary curriculum mechanism jointly estimates round-dependent source reliability and target-specific historical update agreement to regulate cross-client contributions. Together, these components construct personalized aggregation bridges that favor relevant and stable clients while suppressing irrelevant or unreliable updates. At the local side, the personalized bridged model from the previous communication round serves as a teacher to generate complementary-task pseudo-labels. Output- and feature-level cross-task alignment then converts heterogeneous external knowledge into task-compatible auxiliary supervision. Task-aware feature modulation further supports adaptive knowledge absorption and reduces task-induced conflict. Overall, TriEKB-Surg determines which knowledge should be transferred, when it should be trusted, and how it can be incorporated into the target task.

3.3 Evolutionary Knowledge Bridging

3.3.1 Dual-Prompt Compatibility.

Under joint data and task heterogeneity, client relevance varies significantly. Thus, treating all clients equally may mix transferable knowledge with noisy or less relevant updates. Therefore, the first question is *which* a target client should learn from. To achieve this, each client is represented by two complementary prompts: a coarse prompt for broad cross-client relatedness and a fine prompt for refined source selection among semantically closer clients.

For client k, the coarse prompt p_k^c describes the broad task family, such as surgical segmentation or depth estimation, and is used to capture cross-client relevance among all clients. The fine prompt p_k^f further specifies task semantics, such as instrument segmentation or anatomy segmentation, and it is used to refine source selection within the same broad task family. Both prompts are encoded by a frozen text encoder $E_{\text{text}}(\cdot)$ (Radford et al., 2021) and ℓ_2-normalized before the compatibility estimation. This normalization makes pairwise relevance depend mainly on semantic direction rather than embedding magnitude to ensure stable estimation. The process can be formulated as:

$$e_k^c = \frac{E_{\text{text}}(p_k^c)}{\|E_{\text{text}}(p_k^c)\|_2}, \tag{3}$$

where e_k^c is the normalized embedding of the coarse prompt. The same formulation is applied to the fine prompt to get e_k^f.

For a target client i and the other client j, the coarse-level pairwise similarity can be computed as:

$$s_{i,j}^c = \text{sim}(e_i^c, e_j^c), \tag{4}$$

where $\mathrm{sim}(\cdot,\cdot)$ denotes cosine similarity. An analogous computation is applied to the fine prompt embeddings to obtain the fine-level pairwise similarity $s^f_{i,j}$, which captures the fine-grained compatibility between clients within the same broad task family.

Then, we obtain the coarse-level and fine-level pairwise compatibility matrices, denoted by $\mathbf{S}^c = [s^c_{i,j}]^K_{i,j=1}$ and $\mathbf{S}^f = [s^f_{i,j}]^K_{i,j=1}$, respectively. This dual-prompt compatibility construction provides a multi-grained client selection mechanism to enable the server to select clients comprehensively. However, this way only determines *which* to learn from, and it does not indicate *when* a source client is sufficiently reliable to contribute, which motivates the following evolutionary curriculum mechanism.

3.3.2 Evolutionary Curriculum

Beyond task-level semantic compatibility, we introduce an evolutionary curriculum module to model the dynamic reliability of cross-client transfer. A source client can be semantically relevant to a target client, yet provide low-quality updates when local training is unstable. We therefore make the knowledge bridge depend not only on *which* source client is relevant, but also on *when* its knowledge should exert an actual influence.

After communication round r, each client i evaluates its locally updated model on its own validation set and reports a task-dependent performance signal m^r_i. In our implementation, segmentation clients use Dice score, whereas depth clients use $1/(1+\mathrm{MAE})$, with MAE denoting the mean absolute error, so that larger values consistently indicate better reliability. Then, each client uploads not only its model update to the server, but also a validation-derived estimate of how trustworthy that update is at the current round.

To avoid making aggregation overly sensitive to transient validation fluctuations, the server smooths these raw reliability results over rounds. The process is formulated as:

$$q^r_i = \lambda q^{r-1}_i + (1-\lambda) m^r_i, \tag{5}$$

where q^r_i denotes the smoothed reliability estimate of client i at round r. λ is the smoothing coefficient. Hence, the server does not overreact to a single unusually good or bad validation outcome.

To make the contribution rule evolve over communication rounds, we introduce a round-dependent curriculum schedule. Let $\rho^r = \min\left(\frac{r}{R_{\mathrm{warm}}}, 1\right)$ be the curriculum ramp, where R_{warm} is the warm-up horizon. The server then interpolates between a strict early threshold and a relaxed late threshold, yielding the round-dependent contribution threshold τ^r at communication round r:

$$\tau^r = (1-\rho^r)\tau_{\mathrm{early}} + \rho^r \tau_{\mathrm{late}}, \tag{6}$$

where τ_{early} and τ_{late} denote the initial and final contribution thresholds, with $\tau_{\mathrm{early}} > \tau_{\mathrm{late}}$.

However, local reliability alone does not necessarily imply transferability to a heterogeneous target client. We therefore measure the instantaneous update agreement from source client j to target client i at round r using the cosine similarity between their uploaded model updates, i.e., $\kappa^r_{i\leftarrow j} = \mathrm{sim}(\Delta\theta^r_i, \Delta\theta^r_j)$. A larger value indicates stronger agreement between their optimization directions, whereas a negative value suggests potential cross-task conflict. To reduce sensitivity to noisy agreement estimates from a single communication round, the server maintains a pairwise historical update agreement as:

$$\bar{\kappa}^r_{i\leftarrow j} = \mu \bar{\kappa}^{r-1}_{i\leftarrow j} + (1-\mu)\kappa^r_{i\leftarrow j}, \tag{7}$$

where $\bar{\kappa}^r_{i\leftarrow j}$ denotes the historical update agreement from source client j to target client i, and μ is the smoothing coefficient. Combining the smoothed source reliability and the pairwise historical update agreement, the server computes the target-specific contribution strength as:

$$g^r_{i\leftarrow j} = \gamma^r_{\min} + (1-\gamma^r_{\min})\,\sigma\left(\frac{q^r_j + \lambda_\kappa \bar{\kappa}^r_{i\leftarrow j} - \tau^r}{T_{\mathrm{gate}}}\right), \tag{8}$$

where $g^r_{i\leftarrow j}$ denotes the contribution strength from source client j to target client i, $\bar{\kappa}^r_{i\leftarrow j}$ is their historical update agreement, λ_κ controls its influence, and T_{gate} is the temperature parameter. We set $\gamma^r_{\min} = \rho^r \gamma_{\mathrm{late}}$ with $\gamma_{\mathrm{late}} = 0.35$. Early in training, τ^r is high and $\gamma^r_{\min}$ is close to zero, so only reliable and target-compatible clients receive large gates. As training progresses, the threshold relaxes and the minimum contribution floor increases, allowing broader collaboration once models become more stable.

3.3.3 PERSONALIZED BRIDGE AGGREGATION

Given the coarse-level and fine-level pairwise compatibility matrices $\mathbf{S}^c$ and $\mathbf{S}^f$, together with the target-specific curriculum gates $\{g^r_{i\leftarrow j}\}_{j=1}^K$, the server constructs a personalized knowledge bridge for each client. Specifically, the compatibility matrices determine *which* source clients are semantically relevant, whereas the curriculum gates determine *when* their knowledge is reliable and compatible with the target client.

For the client i, the server performs both coarse- and fine-level aggregation. In both cases, the target-specific curriculum gate modulates the aggregation weights. For example, the coarse-level aggregation weight is defined as:

$$\tilde{w}^{\mathrm{c},r}_{i,j} = \frac{g^r_{i\leftarrow j}\exp\big(\mathbf{S}^c_{i,j}\big)}{\sum_{k=1}^K g^r_{i\leftarrow k}\exp\Big(\mathbf{S}^c_{i,k}\Big)+\epsilon}, \tag{9}$$

where $\mathbf{S}^c_{i,j}$ denotes the (i,j)-th entry of $\mathbf{S}^c$, and ϵ is a small constant for numerical stability. The fine-level aggregation follows the same formulation by replacing $\mathbf{S}^c_{i,j}$ with $\mathbf{S}^f_{i,j}$.

Accordingly, the server computes the coarse-level and fine-level transfer terms for client i as:

$$\Delta\theta^r_{i,\mathrm{c}} = \sum_{j=1}^K \tilde{w}^{\mathrm{c},r}_{i,j}\Delta\theta^r_j, \tag{10}$$

where $\Delta\theta^r_j$ denotes the model update uploaded by client j at round r. The fine-level transfer term $\Delta\theta^r_{i,\mathrm{f}}$ is computed in the same manner using the fine-level compatibility matrix.

Finally, the server combines the local update, the fine-level transfer term, and the coarse-level transfer term to form the personalized bridged update:

$$\Delta\theta^r_{i,\mathrm{b}} = \Delta\theta^r_i + \alpha\left(\Delta\theta^r_{i,\mathrm{f}} + \Delta\theta^r_{i,\mathrm{c}}\right), \tag{11}$$

where α denotes the balancing coefficient.

The server then sends the personalized bridged update $\Delta\theta^r_{i,\mathrm{b}}$ back to client i. In this way, the proposed aggregation enables each client to select reliable and compatible source clients at different training stages, thereby achieving evolutionary and personalized model updating.

3.4 CROSS-TASK ALIGNMENT

Semantic relevance and dynamic reliability determine whether knowledge from a source client should be transferred, but they do not resolve the optimization discrepancy caused by heterogeneous task objectives. To address this issue, we propose a cross-task alignment regularization module that encourages the target client to absorb transferred knowledge through complementary task supervision. Instead of enforcing direct parameter-level correction, the transferred knowledge is introduced as a soft auxiliary constraint that guides local optimization.

At communication round r, the personalized model from the previous round $\theta^{r-1}_{\mathrm{b}}$ serves as the teacher, while the current model is treated as the student. The teacher generates task-adaptive auxiliary pseudo-labels: depth supervision for segmentation clients and segmentation supervision for depth clients. We perform alignment at both output and feature levels. Output-level alignment transfers complementary semantic knowledge through task predictions, while feature-level alignment reduces representation discrepancy caused by heterogeneous task objectives. Then, the overall local objective is formulated as:

$$\mathcal{L}^i = \mathcal{L}^i_{\mathrm{task}} + \eta^r_i\left(\mathcal{L}^i_{\mathrm{out}} + \mathcal{L}^i_{\mathrm{feat}}\right), \tag{12}$$

where $\mathcal{L}^i_{\mathrm{task}}$ denotes the main task loss of client i, $\mathcal{L}^r_{i,\mathrm{out}}$ and $\mathcal{L}^r_{i,\mathrm{feat}}$ denote the output- and feature-level alignment losses, respectively. η^r_i is a dynamic balancing factor computed from the relative gradient magnitudes of the primary and auxiliary objectives, preventing transferred knowledge from dominating local optimization.

Beyond alignment, effective knowledge transfer requires task-compatible feature adaptation. Different clients should not process transferred knowledge with identical feature transformations. Therefore, we introduce task-aware feature modulation by combining a dynamic task condition with a

fixed task condition. The dynamic condition provides adaptive modulation during training, while the fixed condition preserves stable task-specific semantics. For each selected shared module, feature modulation is performed as:

$$\tilde{h} = h \odot \left(1 + \beta\, \phi\Big(c_i^{\mathrm{dyn}} + c_i^{\mathrm{fix}}\Big)\right), \tag{13}$$

where h and $\tilde{h}$ denote intermediate features before and after modulation, β controls modulation strength, $\phi(\cdot)$ is the learnable modulation mapping, c_i^{dyn} and c_i^{fix} represent dynamic and fixed task conditions, respectively. $\odot$ denotes element-wise multiplication.

Through cross-task alignment and task-aware modulation, TriEKB-Surg enables task-compatible knowledge transfer and adaptive absorption under heterogeneous FL.

4 Theoretical Analysis

In this section, we provide a theoretical analysis to support our motivation. Let $J_k(\theta)$ denote the local objective of client k evaluated at a common reference parameter θ, and let $\mathbf{v}_k(\theta) = \nabla_\theta J_k(\theta)$ denote its local gradient. Evaluating all objectives at the same parameter space removes the additional heterogeneity caused by different client model architectures, thereby providing a setting favorable to conventional aggregation.

Proposition 1 (Aggregatable Updates Do Not Guarantee Transferable Knowledge) *For a source client j, let $\mathbf{v}_j(\theta) = \nabla J_j(\theta)$. If $\mathbf{v}_j(\theta) \neq \mathbf{0}$, then the source direction $-\mathbf{v}_j(\theta)$ decreases J_j for a sufficiently small step size. However, the same direction is a first-order descent direction for a target objective J_i if and only if $\langle \mathbf{v}_i(\theta), \mathbf{v}_j(\theta) \rangle > 0$.*

Therefore, an update that improves its source objective can be numerically aggregated but may not provide transferable knowledge to a heterogeneous target task. This motivates the need to model cross-client relevance, temporal reliability, and task compatibility rather than relying on unconditional aggregation.

Theorem 1 (Sensitivity of Transfer Utility) *Define the first-order transfer utility from source client j to target client i at round r as: $U_{i\leftarrow j}^r = \big\langle \nabla J_i(\theta_i^r), \nabla J_j(\theta_j^r) \big\rangle$. Under Lipschitz-continuous and bounded gradients, the variation of $U_{i\leftarrow j}^r$ across communication rounds is bounded by the changes in both clients' model states, while its variation across target clients is bounded by the difference between their task gradients. Thus, transfer utility is jointly target-specific and state-dependent.*

Overall, Proposition 1 establishes that an update benefiting its source task is not necessarily transferable to a heterogeneous target task, even when both updates share the same parameter space. Theorem 1 further demonstrates that transfer utility evolves with optimization states and differs across target tasks. These results motivate replacing fixed aggregation rules with dynamic, client-specific knowledge bridging in task-heterogeneous FL. Detailed proofs are provided in the ***Appendix***.

5 Experiments

5.1 Experimental Setup

We construct a heterogeneous multi-task FL setting with five public datasets. Specifically, 1) EndoVis2017 (Allan et al., 2019) focuses on surgical instrument part segmentation, containing 3 classes from nephrectomy procedures. 2) EndoVis2018 (Allan et al., 2020) extends segmentation to the whole surgical scene, covering 11 classes within the same surgical type. 3) AutoLaparo (Wang et al., 2022) provides surgical scene segmentation with 9 classes for hysterectomy procedures. 4) SCARED (Allan et al., 2021) is designed for depth estimation in porcine cadaver abdominal anatomy. 5) StereoMIS (Hayoz et al., 2023) consists of endoscopic videos from in vivo porcine subjects for depth estimation. Following our heterogeneous federation setting, EndoVis2017, EndoVis2018, and AutoLaparo are used as segmentation clients, while SCARED and StereoMIS are used as depth-estimation clients. For evaluation, we report Dice and IoU for segmentation tasks,

Table 1: Quantitative comparison on different clients. The best results are shown in bold.

Method	StereoMIS		SCARED		EndoVis2017		EndoVis2018		AutoLaparo		Δ_m (%) ↑
	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	IoU ↑	Dice ↑	IoU ↑	Dice ↑	IoU ↑	Dice ↑	
Inference	85.037	90.123	92.924	96.353	0.502	0.570	0.507	0.579	0.803	0.868	-151.85
FedAvg	15.216	19.125	21.393	25.758	0.616	0.703	0.633	0.718	0.821	0.877	0.00
FedProx	26.683	32.083	22.083	26.283	0.644	0.726	0.244	0.334	0.827	0.881	-25.42
FedProx*	17.389	20.756	10.435	12.750	0.518	0.598	0.536	0.619	0.765	0.833	0.72
FedAWA	15.420	18.900	**8.910**	**11.120**	0.450	0.535	0.440	0.526	0.784	0.853	-0.02
FUSS	12.000	16.220	11.830	15.160	0.622	0.690	0.654	0.722	0.832	0.889	12.80
FedBN	20.761	24.210	13.560	17.082	0.609	0.678	0.627	0.706	0.822	0.880	0.05
LoRA-Fair	63.111	71.086	22.573	27.137	0.626	0.704	0.635	0.715	0.835	0.889	-59.25
FedNova	11.683	16.406	34.483	39.476	0.567	0.629	0.606	0.677	0.788	0.844	-11.31
FedPer	26.760	37.405	34.911	40.914	0.590	0.660	0.616	0.689	0.883	0.888	-30.17
FedHCA2	11.113	15.159	12.353	15.726	0.6315	0.700	0.648	0.724	0.831	0.886	13.64
MaTFL	11.457	15.599	14.921	17.945	0.634	0.701	0.647	0.718	0.832	0.886	11.10
HNFL	11.249	15.439	13.777	16.934	0.625	0.689	0.644	0.719	0.827	0.883	11.80
Ours	**10.390**	**14.370**	12.516	15.489	**0.658**	**0.729**	**0.660**	**0.726**	**0.837**	**0.892**	**15.76**

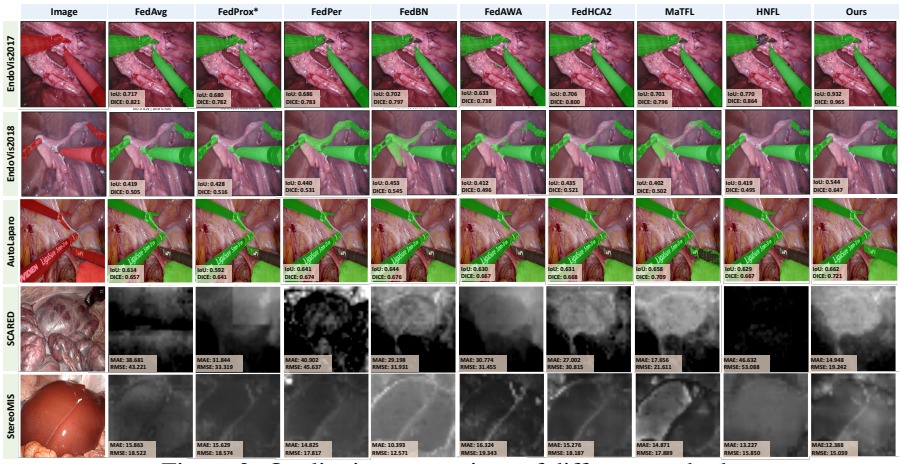


Figure 3: Qualitative comparison of different methods.

and MAE and RMSE for depth estimation. Following Fang et al. (2026a), we use Δ_m to denote the overall relative improvement over the baseline (FedAvg), which can be computed as:

$$\Delta_m = \frac{1}{N}\sum_{i=1}^{N} d_i \frac{M_i - M_i^{\text{FedAvg}}}{M_i^{\text{FedAvg}}} \times 100\%, \tag{14}$$

where N is the number of evaluation metrics, and $d_i = +1$ for higher-is-better metrics and $d_i = -1$.

We employ the SAM2 (Zhu et al., 2024) as the backbone. During inference, we provide only the first frame's bounding box. For the segmentation task, this simulates the surgeon's instruction regarding the region of interest at the start of the surgery; for depth estimation, we utilize a whole-image-size bounding box as the prompt. We freeze the encoder while fine-tuning the memory and decoder layers, and the model is optimized using Adam (Kingma & Ba, 2014) with a learning rate of 1e-4 and 100 communication rounds. All experiments were conducted on an NVIDIA RTX 3090 GPU.

We compare it with multiple advanced methods, including FedAvg (McMahan et al., 2017), FedProx (Li et al., 2020), FedProx (task-specific aggregation), FedNova (Wang et al., 2020), FedBN (Li et al., 2021), LoRA-Fair (Bian et al., 2025), FedAWA (Shi et al., 2025), FUSS (Charalampakis et al., 2026), FedHCA2 (Lu et al., 2024), MaTFL (Cai et al., 2023), and HNFL (Fang et al., 2026a).

5.2 Main Results

Tab. 1 compares different FL methods across two depth-estimation and three segmentation datasets. Our method achieves the best overall performance with a Δ_m of 15.76%, while obtaining the best results on StereoMIS and all three segmentation datasets. Although FedAWA performs best on SCARED, its segmentation performance is notably weaker. The main reason is that it avoids forcing heterogeneous clients toward a single shared optimization direction. Instead, it preserves task-specific representations while transferring only knowledge that is compatible with the target task. This reduces destructive interference between depth estimation and segmentation gradients, while still allowing both tasks to benefit from shared low-level visual structures. As a result, the model achieves a better balance between cross-client knowledge sharing and task-specific adapta-

tion. Moreover, the qualitative results presented in Fig. 3 further demonstrate the effectiveness of our method in producing accurate and visually consistent predictions across heterogeneous tasks.

5.3 Ablation Studies

To evaluate the effectiveness of different components, the ablation study is illustrated in Tab. 2. FedAvg is treated as baseline; "Ours†" and "Ours‡" denote variants of our method without the evolutionary knowledge bridge and the cross-task alignment regularization module, respectively. It can be observed that removing either component results in overall performance degradation, demonstrating that both modules contribute meaningfully to the final performance. Moreover, we observe that removing either p_k^c or p_k^f reduces the improvement, which demonstrates that the two prompts provide complementary guidance for effective knowledge bridging across heterogeneous surgical tasks.

Table 2: Ablation study of the proposed framework.

Method	StereoMIS MAE ↓	SCARED MAE ↓	EndoVis2017 Dice ↑	EndoVis2018 Dice ↑	AutoLaparo Dice ↑	Δ_m (%) ↑
Baseline	15.216	21.393	0.703	0.718	0.877	-
Ours†	11.372	13.990	0.714	0.731	0.885	12.83
Ours‡	11.159	14.450	0.700	0.728	0.888	12.27
Ours w/o p_k^c	11.166	15.392	0.707	0.706	0.890	11.01
Ours w/o p_k^f	11.248	14.816	0.694	0.725	0.889	11.57
Ours	10.390	12.516	0.729	0.726	0.892	15.95

Table 3: The hyperparameter selection of β on Eq (13).

β	StereoMIS MAE ↓	SCARED MAE ↓	EndoVis2017 Dice ↑	EndoVis2018 Dice ↑	AutoLaparo Dice ↑	Δ_m (%) ↑
0.1	11.780	15.120	0.715	0.716	0.889	10.92
0.2	11.805	**11.290**	0.716	0.728	0.889	14.86
0.3	**10.390**	12.516	**0.729**	0.726	**0.892**	**15.95**
0.4	11.137	16.177	0.704	**0.730**	0.884	10.76
0.5	11.105	13.438	0.713	0.717	0.887	13.33
0.6	10.967	13.325	0.682	0.725	0.886	12.92
Avg.	11.197 $\pm_{0.534}$	13.645 $\pm_{1.764}$	0.710 $\pm_{0.016}$	0.724 $\pm_{0.006}$	0.888 $\pm_{0.003}$	13.12 $\pm_{2.08}$

Tab. 3 investigates the sensitivity of our method to the hyperparameter β. Overall, our method remains stable across a broad range of values. When β is too small, the transferred knowledge may be insufficient to fully exploit complementary information from heterogeneous clients. In contrast, excessively large values can overemphasize transferred updates and introduce interference between heterogeneous task objectives. Hence, we suggest $\beta = 0.3$ as the default setting.

5.4 Analysis of Knowledge Bridging

Fig. 4 illustrates the inter-client collaboration relationships over communication rounds. These relationships are not static, and the relative importance of different source clients changes continuously as local models evolve. This supports our motivation that the compatibility and utility of transferred knowledge are time-varying. Meanwhile, the collaboration becomes relatively stable during the later training stages, with only minor variations. Hence, our evolutionary strategy gradually stabilizes knowledge transfer once reliable inter-client relationships have been established. Moreover, Fig. 5 presents the performance gains achieved at different stages of federated training. Our method consistently delivers positive improvements across most communication rounds for both depth estimation and segmentation tasks. Although the magnitude of these gains varies as the local models and inter-client collaboration relationships evolve, the benefits persist throughout the training process rather than being confined to a particular stage. Our proposed evolutionary strategy remains effective under different optimization stages.

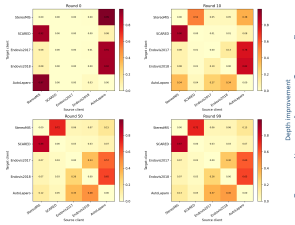


Figure 4: The relationships among clients over communication rounds.

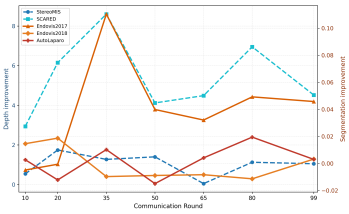


Figure 5: The improvement across communication rounds.

6 Conclusion

In this paper, we reformulated task-heterogeneous federated surgical video understanding as an evolving and client-specific knowledge-transfer problem rather than a static parameter-aggregation process. Based on this perspective, we proposed TriEKB-Surg to jointly determine which knowledge should be transferred, when it should be trusted, and how it should be absorbed. Theoretical analysis and experiments on five surgical datasets demonstrate the effectiveness of the proposed evolutionary knowledge-bridging framework. In future work, we will extend this framework to more diverse surgical video-language understanding tasks, including video captioning and temporal grounding.

AI USE STATEMENT

In this work, we used generative AI tools only for limited writing assistance, including language polishing, grammar correction, and stylistic refinement. We did not use generative AI tools for technical content, method design, experiments, or result analysis. All AI-assisted text was reviewed and verified by the authors, who take full responsibility for the final content.

REFERENCES

Max Allan, Alex Shvets, Thomas Kurmann, Zichen Zhang, Rahul Duggal, Yun-Hsuan Su, Nicola Rieke, Iro Laina, Niveditha Kalavakonda, Sebastian Bodenstedt, et al. 2017 robotic instrument segmentation challenge. *arXiv preprint arXiv:1902.06426*, 2019.

Max Allan, Satoshi Kondo, Sebastian Bodenstedt, Stefan Leger, Rahim Kadkhodamohammadi, Imanol Luengo, Felix Fuentes, Evangello Flouty, Ahmed Mohammed, Marius Pedersen, et al. 2018 robotic scene segmentation challenge. *arXiv preprint arXiv:2001.11190*, 2020.

Max Allan, Jonathan Mcleod, Congcong Wang, Jean Claude Rosenthal, Zhenglei Hu, Niklas Gard, Peter Eisert, Ke Xue Fu, Trevor Zeffiro, Wenyao Xia, et al. Stereo correspondence and reconstruction of endoscopic data challenge. *arXiv preprint arXiv:2101.01133*, 2021.

Manoj Ghuhan Arivazhagan, Vinay Aggarwal, Aaditya Kumar Singh, and Sunav Choudhary. Federated learning with personalization layers. *arXiv preprint arXiv:1912.00818*, 2019.

Jieming Bian, Lei Wang, Letian Zhang, and Jie Xu. Lora-fair: Federated lora fine-tuning with aggregation and initialization refinement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3737–3746, 2025.

Ruisi Cai, Xiaohan Chen, Shiwei Liu, Jayanth Srinivasa, Myungjin Lee, Ramana Kompella, and Zhangyang Wang. Many-task federated learning: A new problem setting and a simple baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5037–5045, 2023.

Evangelos Charalampakis et al. Federated unsupervised semantic segmentation. *Neurocomputing*, pp. 132799, 2026.

Zheng Fang, Ziwei Niu, Ziyue Wang, Zhu Zhuo, Haofeng Liu, Shuyang Qian, Jun Xia, and Yueming Jin. Surgfed: Language-guided multi-task federated learning for surgical video understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6630–6639, 2026a.

Zheng Fang, Xiaoming Qi, Chun-Mei Feng, Jialun Pei, Weixin Si, and Yueming Jin. Spatio-temporal representation decoupling and enhancement for federated instrument segmentation in surgical videos. *IEEE Transactions on Medical Imaging*, 2026b.

Michel Hayoz, Christopher Hahne, Mathias Gallardo, Daniel Candinas, Thomas Kurmann, Maximilian Allan, and Raphael Sznitman. Learning how to robustly estimate camera pose in endoscopic videos. *International journal of computer assisted radiology and surgery*, 18(7):1185–1192, 2023.

Ufaq Khan, Umair Nawaz, Adnan Qayyum, Shazad Ashraf, Yutong Xie, Muhammad Haris Khan, Muhammad Bilal, and Junaid Qadir. Surgical scene understanding in the era of foundation ai models: A comprehensive review. *arXiv preprint arXiv:2502.14886*, 2025.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4015–4026, 2023.

Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. *Proceedings of Machine learning and systems*, 2:429–450, 2020.

Xiaoxiao Li, Meirui JIANG, Xiaofei Zhang, Michael Kamp, and Qi Dou. Fedbn: Federated learning on non-iid features via local batch normalization. In *International Conference on Learning Representations*, 2021.

Xue Li, Wenxin Chen, Xingguang Duan, Xiaoyi Gu, and Changsheng Li. Collaborative surgical instrument segmentation for monocular depth estimation in minimally invasive surgery. *Medical Image Analysis*, pp. 103765, 2025.

Haofeng Liu, Erli Zhang, Junde Wu, Mingxuan Hong, and Yueming Jin. Surgical sam 2: Real-time segment anything in surgical video by efficient frame pruning. *arXiv preprint arXiv:2408.07931*, 2024.

Ange Lou, Yamin Li, Yike Zhang, and Jack Noble. Surgical depth anything: depth estimation for surgical scenes using foundation models. In *Medical Imaging 2025: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 13408, pp. 77–82. SPIE, 2025.

Yuxiang Lu, Suizhi Huang, Yuwen Yang, Shalayiding Sirejiding, Yue Ding, and Hongtao Lu. Fedhca2: Towards hetero-client federated multi-task learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5599–5609, 2024.

Lena Maier-Hein, Matthias Eisenmann, Duygu Sarikaya, Keno März, Toby Collins, Anand Malpani, Johannes Fallert, Hubertus Feussner, Stamatia Giannarou, Pietro Mascagni, et al. Surgical data science–from concepts toward clinical translation. *Medical image analysis*, 76:102306, 2022.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pp. 1273–1282. Pmlr, 2017.

Adithyavairavan Murali, Animesh Garg, Sanjay Krishnan, Florian T Pokorny, Pieter Abbeel, Trevor Darrell, and Ken Goldberg. Tsc-dl: Unsupervised trajectory segmentation of multi-modal surgical demonstrations with deep learning. In *2016 IEEE international conference on robotics and automation (ICRA)*, pp. 4150–4157. IEEE, 2016.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, et al. Sam 2: Segment anything in images and videos. In *International Conference on Learning Representations*, volume 2025, pp. 28085–28128, 2025.

Jan Sellner, Silvia Seidlitz, Alexander Studier-Fischer, Alessandro Motta, Berkin Özdemir, Beat Peter Müller-Stich, Felix Nickel, and Lena Maier-Hein. Semantic segmentation of surgical hyperspectral images under geometric domain shifts. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 618–627. Springer, 2023.

Changlong Shi, He Zhao, Bingjie Zhang, Mingyuan Zhou, Dandan Guo, and Yi Chang. Fedawa: adaptive optimization of aggregation weights in federated learning using client vectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 30651–30660, 2025.

Mengyu Sun, Ziyuan Yang, Yongqiang Huang, Hui Yu, Yingyu Chen, Shuren Qi, Andrew Beng Jin Teoh, and Yi Zhang. Federated learning for large models in medical imaging: a comprehensive review. *arXiv preprint arXiv:2508.20414*, 2025.

Rong Tao, Baoru Huang, Xiaoyang Zou, and Guoyan Zheng. Svt-sde: Spatiotemporal vision transformers-based self-supervised depth estimation in stereoscopic surgical videos. *IEEE Transactions on Medical Robotics and Bionics*, 5(1):42–53, 2023.

Jianyu Wang, Qinghua Liu, Hao Liang, Gauri Joshi, and H Vincent Poor. Tackling the objective inconsistency problem in heterogeneous federated optimization. *Advances in neural information processing systems*, 33:7611–7623, 2020.

Ziyi Wang, Bo Lu, Yonghao Long, Fangxun Zhong, Tak-Hong Cheung, Qi Dou, and Yunhui Liu. Autolaparo: A new dataset of integrated multi-tasks for image-guided surgical automation in laparoscopic hysterectomy. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 486–496. Springer, 2022.

Jialang Xu, Jiacheng Wang, Lequan Yu, Danail Stoyanov, Yueming Jin, and Evangelos B Mazomenos. Personalizing federated instrument segmentation with visual trait priors in robotic surgery. *IEEE Transactions on Biomedical Engineering*, 72(6):1886–1896, 2025.

Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything: Unleashing the power of large-scale unlabeled data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10371–10381, 2024.

Ziyuan Yang, Yingyu Chen, Zhiwen Wang, Hongming Shan, Yang Chen, and Yi Zhang. Patient-level anatomy meets scanning-level physics: Personalized federated low-dose ct denoising empowered by large language model. In *Proceedings of the computer vision and pattern recognition conference*, pp. 5154–5163, 2025a.

Ziyuan Yang, Wenjun Xia, Zexin Lu, Yingyu Chen, Xiaoxiao Li, and Yi Zhang. Hypernetwork-based physics-driven personalized federated learning for ct imaging. *IEEE transactions on neural networks and learning systems*, 36(2):3136–3150, 2025b.

Wenxi Yue, Jing Zhang, Kun Hu, Yong Xia, Jiebo Luo, and Zhiyong Wang. Surgicalsam: Efficient class promptable surgical instrument segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 6890–6898, 2024.

Jiayuan Zhu, Abdullah Hamdi, Yunli Qi, Yueming Jin, and Junde Wu. Medical sam 2: Segment medical images as video via segment anything model 2. *arXiv preprint arXiv:2408.00874*, 2024.

APPENDIX

A THEORETICAL PROOF

The key rationale of our analysis is that the optimization direction of the source client is not necessarily beneficial to a heterogeneous target client. Although source updates are algebraically compatible and can be aggregated in the parameter space, their actual transfer effect is determined by the alignment between the source and target task gradients. Moreover, this alignment is not static: it varies with the optimization states of both clients and differs across target tasks. To support these statements, we provide the following theoretical analysis.

A.1 AGGREGATABLE UPDATES DO NOT GUARANTEE TRANSFERABLE KNOWLEDGE

The models maintained by different clients share the same architecture and therefore have parameter vectors of identical dimensionality. This structural compatibility makes their local updates algebraically aggregatable in the parameter space. However, such compatibility does not guarantee that an update beneficial to one source task will also improve a heterogeneous target task.

Specifically, for each client i, let $J_i(\theta)$ denote its local objective function, and define the corresponding gradient at the parameter θ of the model as:

$$\mathbf{v}_i(\theta) = \nabla J_i(\theta), \tag{A1}$$

where $\mathbf{v}_i(\theta)$ represents the local optimization direction induced by the optimization objective of client i at θ.

Consider a source client j with $\mathbf{v}_j(\theta) \neq \mathbf{0}$. Since J_j is differentiable, the first-order expansion of J_j along the source update direction $-\mathbf{v}_j(\theta)$ is

$$\begin{aligned} J_j\big(\theta - \eta \mathbf{v}_j(\theta)\big) &= J_j(\theta) - \eta \left\langle \nabla J_j(\theta), \mathbf{v}_j(\theta) \right\rangle + o(\eta) \\ &= J_j(\theta) - \eta \|\mathbf{v}_j(\theta)\|_2^2 + o(\eta). \end{aligned} \tag{A2}$$

Because $\mathbf{v}_j(\theta) \neq \mathbf{0}$, the first-order term is strictly negative. Therefore, for a sufficiently small step size $\eta > 0$, we can obtain:

$$J_j\big(\theta - \eta \mathbf{v}_j(\theta)\big) < J_j(\theta). \tag{A3}$$

Hence, $-\mathbf{v}_j(\theta)$ is always a local descent direction for the source objective J_j.

For a heterogeneous target client i, applying the same source direction gives

$$\begin{aligned} J_i\big(\theta - \eta \mathbf{v}_j(\theta)\big) &= J_i(\theta) - \eta \left\langle \nabla J_i(\theta), \mathbf{v}_j(\theta) \right\rangle + o(\eta) \\ &= J_i(\theta) - \eta \left\langle \mathbf{v}_i(\theta), \mathbf{v}_j(\theta) \right\rangle + o(\eta). \end{aligned} \tag{A4}$$

Therefore, the first-order change of the target objective is negative if and only if

$$\left\langle \mathbf{v}_i(\theta), \mathbf{v}_j(\theta) \right\rangle > 0. \tag{A5}$$

This establishes a necessary and sufficient first-order condition for cross-task transferability. When the source and target gradients are positively aligned, the source direction decreases both objectives locally. When their inner product is zero, the source update provides no first-order benefit to the target. When their inner product is negative, the source update increases the target objective to first order and consequently induces negative transfer.

Importantly, gradient misalignment is not merely a pathological or rare case in heterogeneous federated learning. In practice, it is likely to occur because different clients may differ in two fundamental aspects: their optimization objectives and their local training data. Objective heterogeneity causes clients to optimize distinct task losses, which can naturally induce different or even conflicting descent directions. Meanwhile, data heterogeneity changes the empirical distributions on which these objectives are evaluated, further altering the magnitude and orientation of the resulting gradients. The joint effect of task-level and data-level heterogeneity therefore makes positive gradient alignment across clients difficult to guarantee.

Thus, an update can be valid and beneficial for its source task while being harmful to another target task. In other words, ***parameter aggregatability does not imply knowledge transferability***.

A.2 Sensitivity of Transfer Utility

Beyond the target-specific nature of cross-task transfer, the training process itself is evolutionary: client models, local gradients, and inter-client relationships continuously change across communication rounds. Consequently, the utility of transferring knowledge from one client to another cannot be regarded as static. To substantiate this observation, we further analyze the sensitivity of transfer utility in this section.

To characterize the dynamic relationship between clients, we define the first-order transfer utility from source client j to target client i at communication round r as

$$U_{i \leftarrow j}^{r} = \left\langle \nabla J_i(\theta_i^r), \nabla J_j(\theta_j^r) \right\rangle. \tag{A6}$$

A positive value indicates that the source gradient is locally compatible with the target objective, whereas a negative value indicates a first-order conflict. Importantly, this utility depends on the current model states of both clients.

Assume that the gradients of J_i and J_j are Lipschitz continuous, then we can get:

$$\|\nabla J_i(\theta) - \nabla J_i(\theta')\|_2 \leq L_i \|\theta - \theta'\|_2, \tag{A7}$$

and

$$\|\nabla J_j(\theta) - \nabla J_j(\theta')\|_2 \leq L_j \|\theta - \theta'\|_2. \tag{A8}$$

We further assume bounded gradient norms as $\|\nabla J_i(\theta)\|_2 \leq G_i$ and $\|\nabla J_j(\theta)\|_2 \leq G_j$.

Then, the variation of transfer utility between two consecutive rounds can be written as

$$U^r_{i\leftarrow j} - U^{r-1}_{i\leftarrow j} = \left\langle \nabla J_i(\theta^r_i), \nabla J_j(\theta^r_j) \right\rangle - \left\langle \nabla J_i(\theta^{r-1}_i), \nabla J_j(\theta^{r-1}_j) \right\rangle . \tag{A9}$$

Adding and subtracting $\langle \nabla J_i(\theta^{r-1}_i), \nabla J_j(\theta^r_j) \rangle$ yields

$$U^r_{i\leftarrow j} - U^{r-1}_{i\leftarrow j} = \left\langle \nabla J_i(\theta^r_i) - \nabla J_i(\theta^{r-1}_i), \nabla J_j(\theta^r_j) \right\rangle + \left\langle \nabla J_i(\theta^{r-1}_i), \nabla J_j(\theta^r_j) - \nabla J_j(\theta^{r-1}_j) \right\rangle . \tag{A10}$$

By the triangle inequality and the Cauchy–Schwarz inequality,

$$\left|U^r_{i\leftarrow j} - U^{r-1}_{i\leftarrow j}\right| \le \left\|\nabla J_i(\theta^r_i) - \nabla J_i(\theta^{r-1}_i)\right\|_2 \left\|\nabla J_j(\theta^r_j)\right\|_2 + \left\|\nabla J_i(\theta^{r-1}_i)\right\|_2 \left\|\nabla J_j(\theta^r_j) - \nabla J_j(\theta^{r-1}_j)\right\|_2 . \tag{A11}$$

Using the Lipschitz-continuity and bounded-gradient assumptions, we obtain

$$\left|U^r_{i\leftarrow j} - U^{r-1}_{i\leftarrow j}\right| \le G_j L_i \|\theta^r_i - \theta^{r-1}_i\|_2 + G_i L_j \|\theta^r_j - \theta^{r-1}_j\|_2 . \tag{A12}$$

This bound shows that transfer utility changes whenever either the source or the target model state changes. Even for a fixed source–target pair, an update that is useful in one communication round may become less useful or even harmful in a later round. Therefore, transferability is inherently state-dependent and cannot be reliably characterized by a static aggregation rule.

A.3 Transfer Utility Across Heterogeneous Targets

We next consider the same source client j and two target clients i and k at round r. Their transfer utilities are

$$U^r_{i\leftarrow j} = \left\langle \nabla J_i(\theta^r_i), \nabla J_j(\theta^r_j) \right\rangle , \tag{A13}$$

and

$$U^r_{k\leftarrow j} = \left\langle \nabla J_k(\theta^r_k), \nabla J_j(\theta^r_j) \right\rangle . \tag{A14}$$

Their difference satisfies:

$$U^r_{i\leftarrow j} - U^r_{k\leftarrow j} = \left\langle \nabla J_i(\theta^r_i) - \nabla J_k(\theta^r_k), \nabla J_j(\theta^r_j) \right\rangle . \tag{A15}$$

Applying the Cauchy–Schwarz inequality gives

$$\left|U^r_{i\leftarrow j} - U^r_{k\leftarrow j}\right| \le \left\|\nabla J_i(\theta^r_i) - \nabla J_k(\theta^r_k)\right\|_2 \left\|\nabla J_j(\theta^r_j)\right\|_2 \le G_j \left\|\nabla J_i(\theta^r_i) - \nabla J_k(\theta^r_k)\right\|_2 . \tag{A16}$$

This result shows that the usefulness of the same source update differs across target clients according to the discrepancy between their task gradients. A source client may be positively aligned with one target while conflicting with another. Consequently, source quality alone is insufficient to determine transferability.

A.4 Implications for Dynamic and Personalized Aggregation

The above theoretical analysis reveal that heterogeneous cross-client transfer cannot be adequately characterized by a single, globally fixed notion of source quality. Although updates from different clients are structurally compatible and can be aggregated in a shared parameter space, their actual utility depends on both the receiving client and the current stage of optimization. The same source update may be beneficial to one target client but ineffective or harmful to another, and its transfer effect may further change as the source and target models evolve over successive communication rounds. These observations jointly establish two fundamental properties of heterogeneous cross-client transfer:

1) Transfer utility is *target-specific*. The same source update can produce different effects on different clients because their task objectives induce different gradient directions.

2) Transfer utility is *state-dependent*. The compatibility between a source and a target changes over communication rounds as their local model states evolve.

Therefore, a reliable aggregation mechanism should not treat client contributions as globally or permanently beneficial. Instead, it should evaluate source knowledge conditionally with respect to both the target client and the current communication state. This provides the theoretical motivation for dynamic personalized aggregation: effective knowledge transfer requires determining not only which source is informative, but also for which target and at which stage of training that source should be trusted.

Nevertheless, dynamic personalized aggregation can only mitigate harmful transfer at the aggregation level; it cannot fundamentally eliminate the discrepancy among clients' local optimization directions. Because different clients still optimize distinct objectives J_i, their gradients may remain misaligned even after source selection and adaptive weighting. Therefore, aggregation-side regulation should be complemented by local objective alignment, which explicitly reduces the discrepancy among heterogeneous task objectives and encourages transferred knowledge to be absorbed in a target-compatible manner. These theoretical analyses motivate us to propose the Triadic Knowledge Evolutionary Bridging framework, which systematically addresses three fundamental questions in heterogeneous cross-client transfer: *which* source clients should be selected, *when* their knowledge should be trusted, and *how* the transferred knowledge should be effectively absorbed by each target client.